

The Impact of Subsampling on MODIS Level-3 Statistics of Cloud Optical Thickness and Effective Radius

Lazaros Oreopoulos

Abstract—The Moderate Resolution Imaging Spectroradiometer (MODIS) Level-3 optical thickness and effective radius cloud product is a gridded $1^\circ \times 1^\circ$ dataset that is derived from aggregation and subsampling of every fifth pixel, along both spatial directions, of Level-2 orbital swath data (Level-2 granules). The present study examines the impact of this subsampling on the mean, standard deviation, and inhomogeneity parameter statistics of optical thickness and effective radius. The methodology is simple and consists of estimating mean errors for a large collection of Terra and Aqua Level-2 granules by taking the difference of the statistics at the original and subsampled resolutions. It is shown that the Level-3 subsampling does not affect the various quantities investigated to the same degree, with second-order moments suffering greater subsampling errors, as expected. Mean errors drop dramatically when averages over a sufficient number of regions (e.g., monthly and/or zonal averages) are taken, pointing to a dominance of errors that are of random nature. When histograms built from subsampled data with the same binning rules as in the Level-3 dataset are used to reconstruct the quantities of interest, the mean errors do not deteriorate significantly. The results in this paper provide guidance to users of MODIS Level-3 optical thickness and effective radius cloud products on the range of errors due to subsampling they should expect and perhaps account for, in scientific work with this dataset. In general, subsampling errors should not be a serious concern when moderate temporal (e.g., monthly) and/or spatial (e.g., zonal) averaging is performed.

Index Terms—Aqua, clouds, effective radius, inhomogeneous media, optical thickness, remote sensing, sampling methods, satellite applications, Terra.

I. INTRODUCTION

IN ORDER to study the global distribution of cloud properties and the main features of their monthly, seasonal, and diurnal evolution, in other words, in order to examine cloud climatology, a gridded set of spatially averaged cloud retrievals would be the most convenient. Such a product is provided for the Moderate Resolution Imaging Spectrometer (MODIS) instrument aboard the Earth Observing System (EOS) Terra and Aqua platforms as Level-3 MOD08 (Terra) and MYD08 (Aqua) datasets [1]. There are actually three Level-3 MODIS cloud products available for each platform. Statistics are summarized over a $1^\circ \times 1^\circ$ global grid for daily (D3), eight-day (E3), and monthly

(M3) time scales. Each of the Level-3 products contain statistics generated from the Level-2 (Orbital Swath) products. Statistics for a given derived quantity or science dataset (SDS) might include: simple (mean, minimum, maximum, standard deviation) statistics; parameters of normal and lognormal distributions; fraction of pixels that satisfy some condition (e.g., cloudy, clear); histograms of the quantity within each gridpoint; histograms of the confidence placed in the retrieved quantity; histograms and/or regressions derived from comparing one science parameter to another; statistics computed for a subset that satisfies some condition [1]. All these statistics are computed by subsampling pixel-level values of 1-km nadir resolution every fifth pixel, along both spatial directions, since the geolocation internal to the MOD06 (Level-2) cloud product is 5 km [1]. Thus, cloud optical thickness or effective radius statistics for an overcast $1^\circ \times 1^\circ$ gridpoint around the equator that contains pixels observed at near-nadir view angles come from about ~ 480 pixels instead of the $\sim 12\,000$ 1-km pixels that are originally contained within the gridpoint. The subject of this study is to examine whether the process of subsampling has distorting effects on several Level-3 SDSs and other quantities of interest derived from them. This is obviously an important issue for current and future users of the Level-3 cloud dataset who intend to compare MODIS cloud climatologies with those from other sources.

The outline of the paper is as follows. First, Section II presents the dataset used to examine the subsampling effects, the SDSs, and other quantities of interest, and discusses the methodology for analyzing the subsampling errors. In Section III, results for optical thickness errors are presented, while Section IV is devoted to effective radius errors. Section V examines whether the findings in Sections III and IV are affected when the quantities of interest are derived from histograms built following Level-3 binning rules for optical thickness and effective radius. The final section consists of an overview discussion on our findings and their implications for users of MODIS Level-3 cloud climatologies.

II. DATASET AND METHODOLOGY

The dataset consists of 300 Level-2 granules, each made of 2030 pixels along track and 1354 pixels across track, obtained for various post-2000 November months for both Terra (200 granules) and Aqua (100 granules). The granules are largely confined within 20°N to 60°N (with most granule centers falling within 20°N and 50°N), but cover the full meridional

Manuscript received June 18, 2004; revised October 29, 2004. This work was supported by the National Aeronautics and Space Administration under Contact NAG5-11631.

The author is with the University of Maryland Baltimore County, Baltimore, MD 21250 USA (e-mail: lazaros@climate.gsfc.nasa.gov).

Digital Object Identifier 10.1109/TGRS.2004.841247

range. They contain a wide variety of cloud types of different phase, cloud fraction, thickness, and degree of inhomogeneity. For those pixels identified as cloudy from the cloud masking algorithm [2], the cloud phase is determined (“liquid,” “ice,” “undetermined”), and subsequently, cloud optical thickness τ , and cloud effective radius r_{eff} (ratio of the third to the second moment of the cloud particle radius distribution) are retrieved (among others) [3]. The retrievals used here come from the 0.65- μm (over land) and 0.86- μm (over ocean) bands that are the most sensitive to changes in cloud optical thickness, in conjunction with the 2.1- μm band, which is most sensitive to changes in cloud particle size [3]. The resolution of these retrievals (pixel size) is 1 km for near-nadir observations, but decreases gradually (pixel size increases) with view angle [4]. In this study, the pixel-by-pixel phase determination will be largely ignored, since it is not an essential factor in subsampling error estimates, as will become evident later. The only instance where phase enters the discussion is in Section V where, due to different histogram binning rules for the two phases, *all* cloudy pixels are assumed to be of one or the other phase.

In the Level-3 dataset, the statistics of each $1^\circ \times 1^\circ$ gridpoint (near the equator) are derived by subsampling and aggregating every fifth of approximately 110×110 pixels near-nadir (1-km resolution). However, the available number of pixels to be subsampled approximately decreases with the cosine of latitude as one moves poleward. For example, at $\sim 83^\circ$, each $1^\circ \times 1^\circ$ gridpoint is made of ~ 1600 1-km pixels. At the same time, for observations close to the edge of the orbital swath, where pixel expansion occurs, fewer pixels are needed to cover a $1^\circ \times 1^\circ$ geographic area (an effect somewhat compensated by accompanying pixel overlap [4]). Thus, the number of pixels used to construct the Level-3 statistics can potentially become quite small, especially when only a fraction of the gridpoint is cloudy (as is often the case). The impact of the varying number of pixels used to construct Level-3 statistics has to be, therefore, taken into account in the analysis.

The approach used in this work is the following. The granules are divided into $110 \times 110, 100 \times 100 \dots 40 \times 40$ pixel regions (i.e., eight regions sizes). Since one of the main goals is to examine the effects of subsampling on the cloud optical thickness inhomogeneity climatology presented in the paper by Oreopoulos and Cahalan [5], only regions with cloud fraction (fraction of pixels with nonzero optical thickness) greater than 0.1 are considered, as in their work. For each of these regions (e.g., $\sim 53\,000$ regions of 110×110 pixels), and for the optical thickness part of the analysis, cloud fraction (CF), spatial mean of optical thickness $\bar{\tau}$, standard deviation of optical thickness σ_τ , and the inhomogeneity parameters

$$\nu_{\text{MOM}} = \left(\frac{\bar{\tau}}{\sigma_\tau} \right)^2 \quad (1a)$$

$$\nu_{\text{MLE}} = \frac{1 + \sqrt{1 + 4y/3}}{4y} \quad (1b)$$

$$\chi = \frac{e^{\ln \bar{\tau}}}{\bar{\tau}}, \quad 0 < \chi \leq 1 \quad (1c)$$

are calculated. In (1b), $y = \ln \bar{\tau} - \overline{\ln \tau}$. The first two equations provide two different ways to estimate the shape parameter of a gamma distribution, which has been found to describe well the observed distributions of cloud optical thickness [6], [7]. The first equation is for the method of moments (MOM), and the second is an empirical approximation for the maximum-likelihood estimate (MLE) method, which gives a shape parameter less sensitive to outliers [8]. The third equation is the definition of the inhomogeneity parameter of Cahalan *et al.* [9], which approximates the factor by which $\bar{\tau}$ should be multiplied to recover the mean albedo of a region. For the effective radius part of the analysis, only the mean and standard deviation are calculated.

For both τ and r_{eff} , two methods of calculation are used: 1) all the cloudy pixels within the region are included; and 2) only every fifth pixel along both spatial directions, if it happens to be cloudy, is included. The percentage difference of the values obtained from the above two methods gives the impact of subsampling as a percentage error (positive signifies that subsampling underestimates). Cloud phase is ignored in this procedure, so the means and standard deviations obtained closely correspond to their counterpart SDSs for “combined optical thickness” and “combined effective radius” in the Level-3 MODIS products.

The analysis shown in the following also accounts for the fact that, at most times, we are not interested in the error of a single region, but in the error of an ensemble of regions. For example, in the work by Oreopoulos and Cahalan [5], the authors are interested in the climatology of χ and ν , i.e., monthly, zonal, and global averages of these quantities. The mean error of an ensemble of 30 regions can then be thought of as the mean monthly sampling error for a single $1^\circ \times 1^\circ$ gridpoint. Similarly, the error for an ensemble of 90 regions can be thought of as the mean seasonal sampling error of a single gridpoint, the error of an ensemble of 360 regions as the mean annual sampling error of a single gridpoint or the daily error of a latitude zone, and the error for an ensemble of 10 000 regions ($\sim 30 \times 360$) as the mean monthly subsampling error of a latitude zone. To examine these “climatological” errors, 1000 ensembles of regions are constructed with each ensemble resulting from assembling in a random fashion, a prespecified number of regions (1, 30, 90, 360, 10 000) for each of the eight region sizes, so that 5000 ensembles correspond to each region size, i.e., 1000 consisting of 1 region, 1000 consisting of 30 regions, etc. The distribution of errors for these 40 000 ensembles can then be examined. The random fashion with which the ensembles are built also ensures that regions observed under a variety of viewing angles are combined, thus crudely simulating the fact that over a period of time, each $1^\circ \times 1^\circ$ gridpoint contains pixels viewed under a wide range of scan angles (a similar argument applies for gridpoints that make up a latitude zone). Still, it is not clear that pixel expansion at large view angles systematically increases the subsampling errors. Indeed, regions consisting of the same number of pixels, but viewed at near-nadir and the largest view angles (i.e., larger regions at the edges of the swath) did not exhibit substantially different subsampling errors in test comparisons. This may be because pixel expansion with view angle is accompanied by simultaneous pixel overlap [4], which tends to reduce subsampling effects.

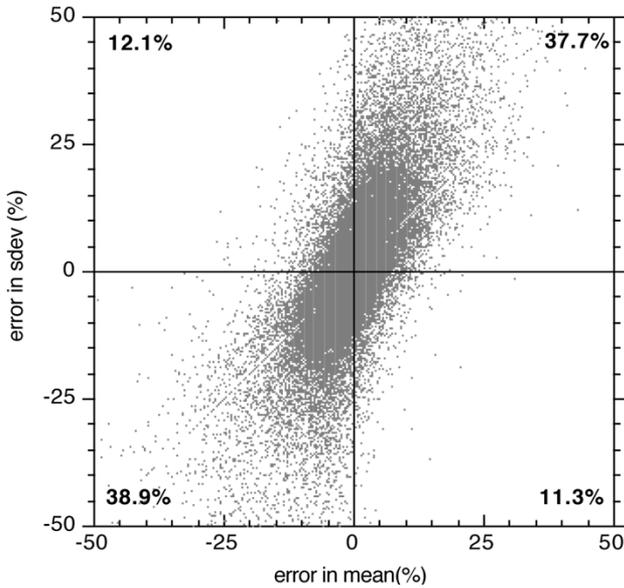


Fig. 1. Subsampling error in the mean and standard deviation of optical thickness (both in percent) for each 110×110 pixel region of our dataset. The numbers in the corners are the percentage of regions with errors falling into the respective quadrants. Note that there were few regions whose errors fell outside the axis limits ($\pm 50\%$) of this plot.

III. OPTICAL THICKNESS ERRORS

Fig. 1 shows subsampling errors (in percent, with the very few falling outside the $\pm 50\%$ range excluded) of $\bar{\tau}$ and σ_{τ} for all ($\sim 53\,700$) 110×110 regions of our dataset. Errors for individual regions are often quite large, although the greatest concentration of points is within the $\pm 20\%$ error bounds. There is about the same number of regions with positive and negative errors in $\bar{\tau}$, and the same applies for σ_{τ} . This is a good indication of the random nature of these errors. For most regions ($\sim 76.6\%$, upper right and lower left quadrants), overestimates in $\bar{\tau}$ by subsampling are accompanied by overestimates in σ_{τ} and vice-versa, but the number of regions where the error is of opposite sign is still substantial. Fig. 2(a) shows a similar graph, but this time for CF and χ . The errors in this case are generally smaller with the densest concentration of points confined within the $\pm 10\%$ error bounds. The number of regions on each quadrant is now distributed more evenly than in the previous figure. Fig. 2(b) shows the percent errors in χ for each region as a function of the cloud fraction of the region before subsampling, and indicates that the distribution of χ errors tightens around smaller values as cloud fraction increases.

Fig. 3 shows the mean error for 110×110 pixel regions as a function of cloud fraction. Each value was obtained by averaging the errors of regions that have cloud fraction within the predetermined 0.1-width bin. Note that the last bin has by far the most values consistent with the well-known U-shape behavior of cloud fraction distributions. The dramatic effect of averaging a large number of random errors is in clear view: the mean errors of ensembles of ~ 5000 regions and above are very small, with the exception of ν_{MOM} at small cloud fractions. The larger impact of sampling on ν_{MOM} compared to the other two inhomogeneity parameters can be easily explained: both χ and ν_{MLE} depend on first moment quantities (the linear mean and the mean

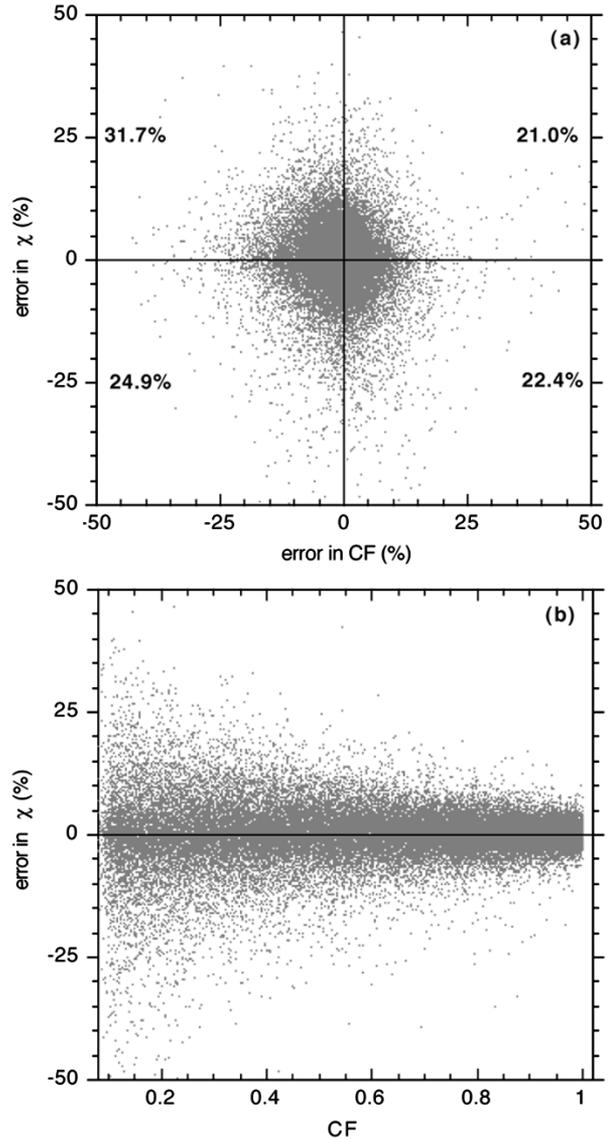


Fig. 2. (a) As in Fig. 1, but for cloud fraction and χ . (b) Subsampling error of χ as a function of the actual (perfectly sampled) CF of each region.

logarithm of optical thickness, the former being simply the ratio $\exp(\ln \bar{\tau})/\bar{\tau}$, and the latter being a function of the difference $\ln \bar{\tau} - \overline{\ln \tau}$, while ν_{MOM} depends on a second moment quantity (σ_{τ}) that is more sensitive to subsampling. Between ν_{MLE} and χ , the latter is less affected by subsampling. There are two reasons for this. First, χ is defined simply as the ratio of two quantities, while ν_{MLE} is a more intricate function of the linear mean and mean logarithm difference (1b) and is, therefore, subject to greater error propagation. Second, χ has an upper bound of 1, by definition, while ν_{MLE} (and, of course, ν_{MOM}) are unbounded. Despite the fact that regions with ν_{MLE} or ν_{MOM} greater than 40 are excluded in the analysis to eliminate contributions from pathological cases, some residual impact remains from regions with large ν_{MOM} , where its value can be easily affected by subsampling. Thus, the unbounded nature of ν_{MOM} is responsible for the apparent paradox that some of the most homogeneous regions can also potentially suffer from the greatest percentage subsampling errors for this particular parameter.

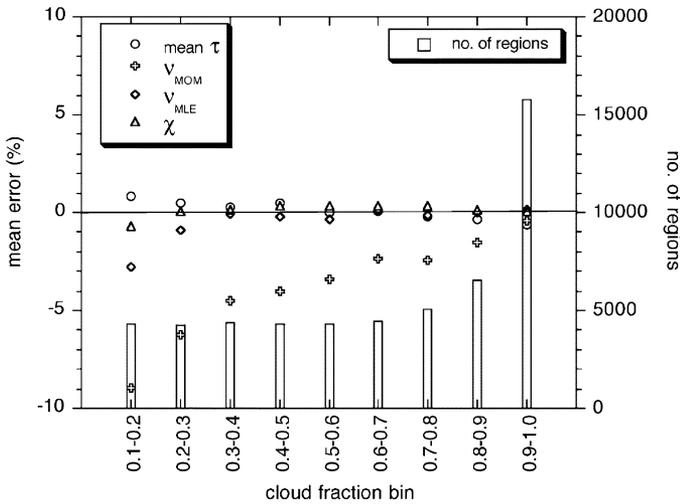


Fig. 3. Mean error (in percent) for various statistics of optical thickness as a function of cloud fraction. The right ordinate shows the number of 110×110 pixel regions with cloud fraction that falls within each 0.1-width bin (regions with cloud fractions less than 0.1 were omitted).

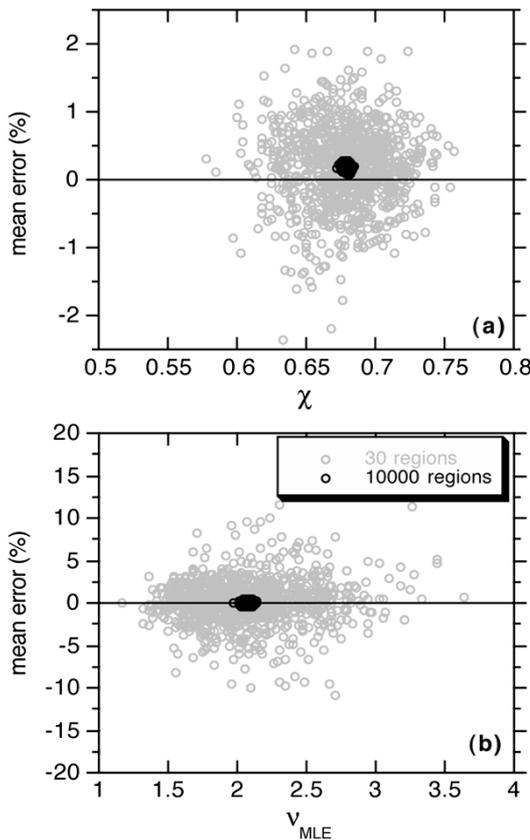


Fig. 4. Mean error (in percent) of (a) χ and (b) ν_{MLE} for each of the 1000 ensembles of 30 and 10000 110×110 pixel regions as a function of the mean value of the ensemble obtained with perfect sampling.

Further evidence of the beneficial effects of averaging errors over a large number of regions is shown in Fig. 4. These percentage errors of χ and ν_{MLE} are for 1000 ensembles each consisting of 30 and 10000 regions. The size of each region in these randomly constructed ensembles is 110×110 pixels. As discussed in Section II, the mean error of an ensemble of 30

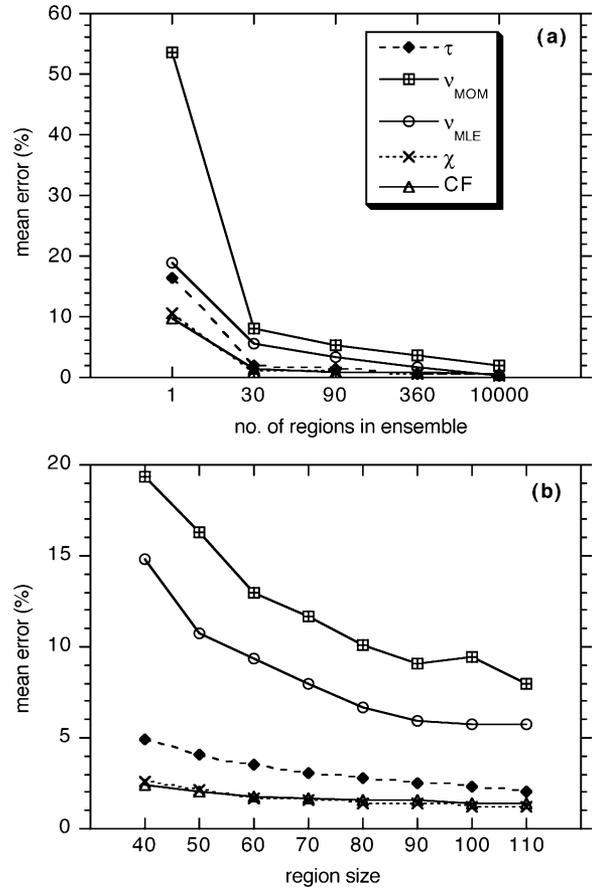


Fig. 5. (a) Subsampling error range (in percent) that contains 95% of the 1000 ensembles each of which is made of the number of 110×110 pixel regions shown in the abscissa. (b) As in (a) panel, but for 1000 ensembles of 30 regions of the size shown in the abscissa.

regions is meant to represent typical monthly average errors of individual gridpoints, while the mean error of an ensemble of 10 000 regions approximates typical monthly-average errors of latitude zones. The mean error of 30-region ensembles almost always stays within $\pm 2\%$ for χ and within $\pm 10\%$ for ν_{MLE} . The mean errors of ensembles consisting of 10 000 regions are much smaller than the 30-region ensembles and cluster within a very small range of values. This clustering is not unexpected since each of the 10 000-region ensembles, even if constructed randomly, contains many common regions with the other ensembles because the population from which it is drawn is only larger by an approximate factor of 5 (there are $\sim 53\,700$ 110×110 regions in the dataset). It is also interesting that the mean errors of 10 000-region ensembles are always positive for χ . This is because of the tight range of χ errors and the fact that a slightly larger number of regions with positive errors is encountered [Fig. 2(a) indicates that 52.7% of regions have positive errors]. On the other hand, because of the wider range of ν_{MLE} errors, there are both positive and negative mean errors for 10 000-region ensembles. The positives dominate due to the larger fraction of positive errors for individual regions ($\sim 55\%$).

Another way to assess the errors of subsampling on optical thickness statistics is shown in Fig. 5. Fig. 5(a) shows the bounds of percentage errors that contain 95% of the 1000 ensembles,

when ensembles consisting of various region numbers (each of 110×110 pixel size) are considered. For example, Fig. 5(a) indicates that 95% (950) of 90-region ensembles have mean errors of ν_{MOM} within $\pm 5.2\%$ (third point of topmost curve). CF and χ have the smallest error bounds that contain 95% of the ensembles, followed by $\bar{\tau}$, ν_{MLE} , and ν_{MOM} . For ensembles consisting of 10 000 regions, the error range that contains 95% of the ensembles is smaller than $\pm 2\%$ for all quantities ($\pm 0.25\%$ for χ !). Fig. 5(b) shows the percentage error range that contains 95% of 30-region ensembles of region size indicated in the abscissa. For example, 95% of 30-region ensembles have ν_{MOM} subsampling errors within $\pm 12.95\%$ when the region size is 60×60 pixels (third point of topmost curve). Because the number of ensembles is kept constant at 1000 for each region size (even if more regions of smaller size actually exist in the dataset), it is not surprising that there is a tendency for the error range that contains 95% of the ensembles to decrease with region size. In other words, subsampling errors are expected to grow for regions consisting of a smaller number of cloudy pixels (i.e., $1^\circ \times 1^\circ$ gridpoints at higher latitudes, gridpoints containing many edge-of-swath pixels, or gridpoints with smaller cloud fractions).

Finally, it must be noted that the errors discussed in this section (as well as in the next two sections) for small region sizes (high latitudes) may in some cases actually be overestimates of the errors resulting from the actual Level-3 processing. The reason is that the number of pixels falling within a $1^\circ \times 1^\circ$ gridpoint each day is greater than that from a single orbital swath. This in turn is because of the overlap of successive Terra and Aqua orbits (except near the equator), which is especially pronounced at high latitudes. Hence, the drastic reduction with latitude of the number of pixels within $1^\circ \times 1^\circ$ gridpoints is to some extent compensated by repeated sampling of the region from multiple satellite passes.

IV. EFFECTIVE RADIUS ERRORS

The analysis in this section follows on the footsteps of that in the previous section. Case in point, Fig. 6 is the counterpart of Fig. 1, i.e., it shows the errors for all 110×110 pixel regions, but this time for the mean and standard deviation of r_{eff} . There are similarities with Fig. 1, such as the rapid decrease in the density of points outside the $\pm 20\%$ error range, but also differences such as the stronger dominance of positive errors for both the mean and the standard deviation. Indeed, only 21.3% of 110×110 regions have negative errors in the mean, and 35.3% have negative errors in the standard deviation. This explains the lack of negative errors when averaging is performed over a larger number of regions, as in Figs. 7 and 8. These show (similar to their counterpart Figs. 3 and 4) the mean error of r_{eff} mean and standard deviation for 110×110 pixel regions falling within different cloud fraction bins (Fig. 7) and the mean errors of 1000 ensembles of 30 or 10 000 110×110 pixel regions as a function of the corresponding perfectly sampled quantity (Fig. 8). Fig. 7 suggests that mean errors of subsampling for mean r_{eff} are slightly greater than those for $\bar{\tau}$, while somewhat unexpectedly the error in standard deviation does not improve with CF; it does, however, improve with region size, as shown in the bottom panel of Fig. 9(b),

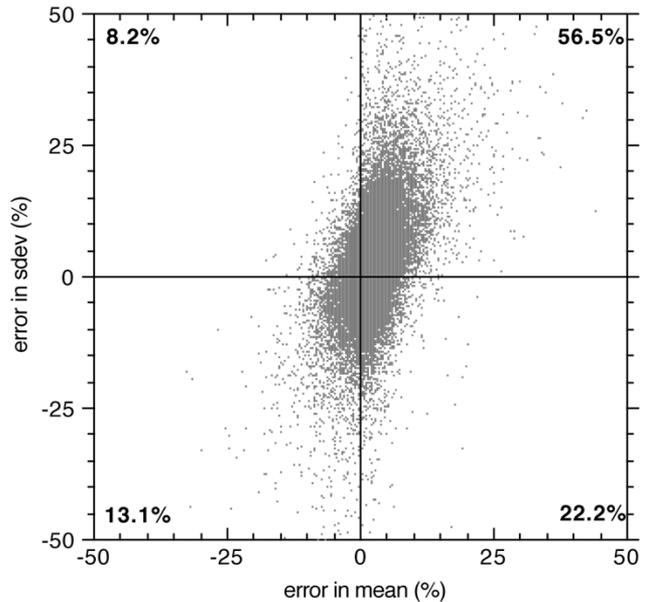


Fig. 6. As in Fig. 1, but for effective radius.

which is analogous to Fig. 5(b), showing the error range containing 95% of the ensembles. Also, there seems to be resistance in reducing the mean errors below 2% even when ensembles consist of 10 000 regions [Figs. 8 and 9(a)].

All the above points to systematic biases in the statistics of r_{eff} when subsampling is performed: apparently, subsampling yields frequent systematic underestimates of both the mean and the standard deviation of r_{eff} , i.e., errors are not always random. This curious phenomenon was further explored by examining effective radius histograms retrieved from perfectly sampled and subsampled data. When the collective (i.e., from all 300 granules) normalized frequency distributions of combined r_{eff} (i.e., both liquid and ice clouds) were plotted with a $1\text{-}\mu\text{m}$ bin resolution (not shown), there were small but noticeable differences between the histograms: a larger normalized frequency at small effective radii and a smaller at large effective radii occurred for the subsampled retrievals. These differences were large enough to result in systematically smaller effective radii for the subsampled data in the majority of regions to which we divide the granules. They also resulted in somewhat narrower histograms for the subsampled data, which explains the tendency for positive subsampling errors in standard deviation. When the histogram analysis was repeated separately for the Terra and Aqua granules, histogram differences appeared only for the Terra platform (incidentally, subsampled and perfectly sampled histograms of optical thickness were virtually indistinguishable for both platforms). When Terra r_{eff} histograms were then constructed separately (not shown) for retrievals corresponding to different pairs of detector elements (the $2.1\text{-}\mu\text{m}$ band has 20 detector elements each of 500-m resolution for a total viewing path of 10 km along track, so for the 1-km effective radius Level-2 product, measurements from two detectors are aggregated), one of the histograms stood out as having characteristics similar to those described above for the ensemble histogram of subsampled data. This histogram was from the detector pair that yielded lines 1, 11, 21, 31, etc.

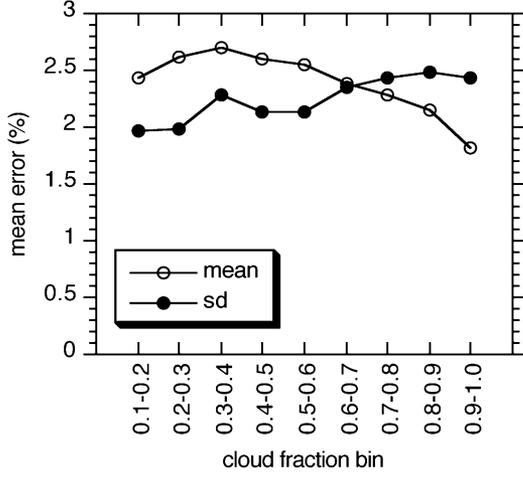


Fig. 7. As in Fig. 3, but for mean and standard deviation of effective radius.

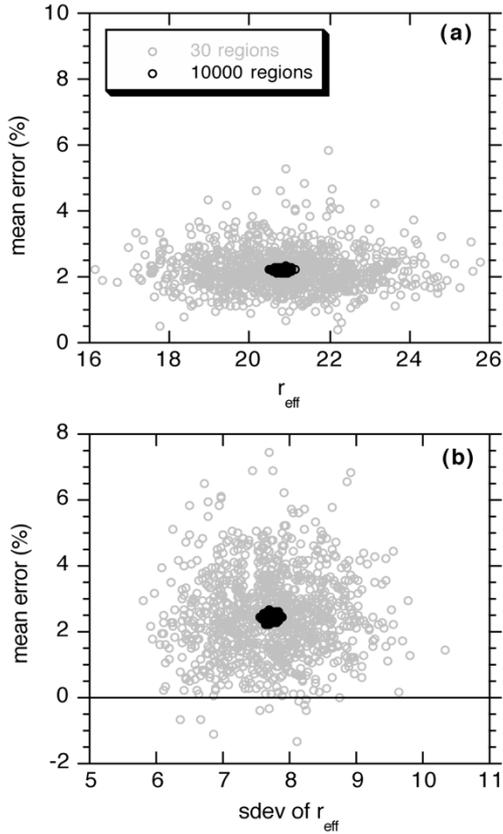


Fig. 8. As in Fig. 4, but for mean and standard deviation of effective radius.

of the granule, which are included in the subsampled dataset of this analysis as well as the actual MODIS Level-3 processing. Thus, bias errors can appear in subsampled Level-3 data if pixel lines with distinct radiative characteristics (and therefore distinct retrieved values) from the other lines are systematically selected by the subsampling algorithm. This is exactly what occurred in this case, and while the bias errors were small in magnitude, they were still easily detected by the subsampling analysis.

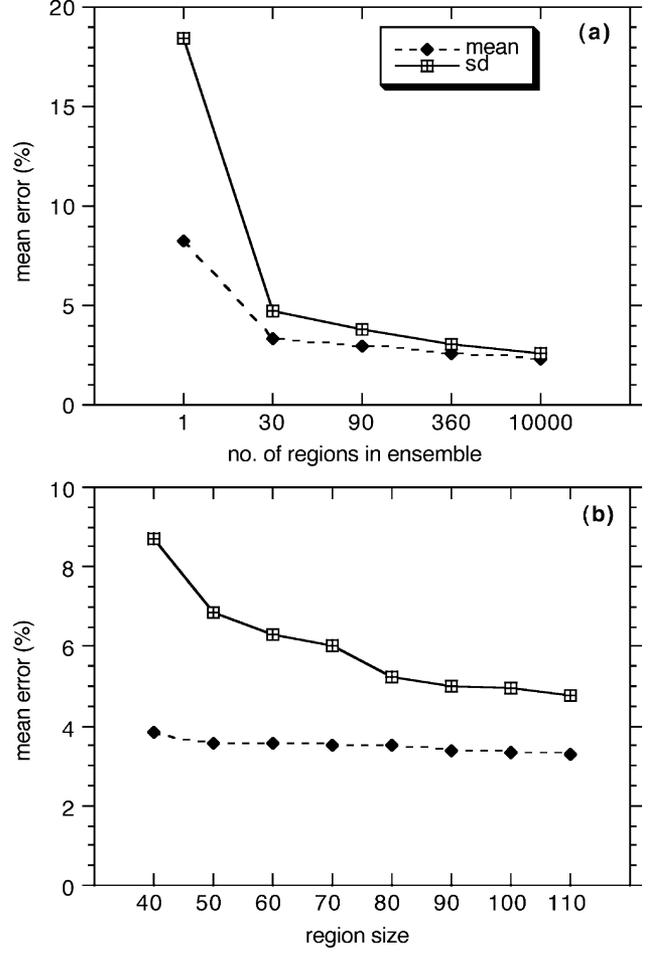


Fig. 9. As in Fig. 5, but for mean and standard deviation of effective radius.

V. ERRORS FROM HISTOGRAMS

The MODIS Level-3 cloud product also includes SDSs that represent histograms of τ and r_{eff} . These are also constructed from subsampled data. Although the statistical quantities and parameters examined here are either given directly as distinct SDS products ($\bar{\tau}$, σ_{τ} , $\overline{\ln \tau}$) or can be trivially derived from them using (1) (χ , ν_{MOM} , ν_{MLE}), it would be interesting to obtain an assessment of the errors when the same quantities are calculated from the histogram SDSs.

The three moments ($\bar{\tau}$, σ_{τ} , $\overline{\ln \tau}$) that are needed for (1) are derived from the discrete normalized probability distribution $p(\tau)$ built from the histograms for each region (of the eight regions sizes) using values subsampled every fifth pixel as follows:

$$\bar{\tau} = \sum_{i=1}^N \tau_i p(\tau_i) \quad (2a)$$

$$\sigma_{\tau} = \left(\sum_{i=1}^N (\tau_i - \bar{\tau})^2 p(\tau_i) \right)^{1/2} \quad (2b)$$

$$\overline{\ln \tau} = \sum_{i=1}^N \ln \tau_i p(\tau_i). \quad (2c)$$

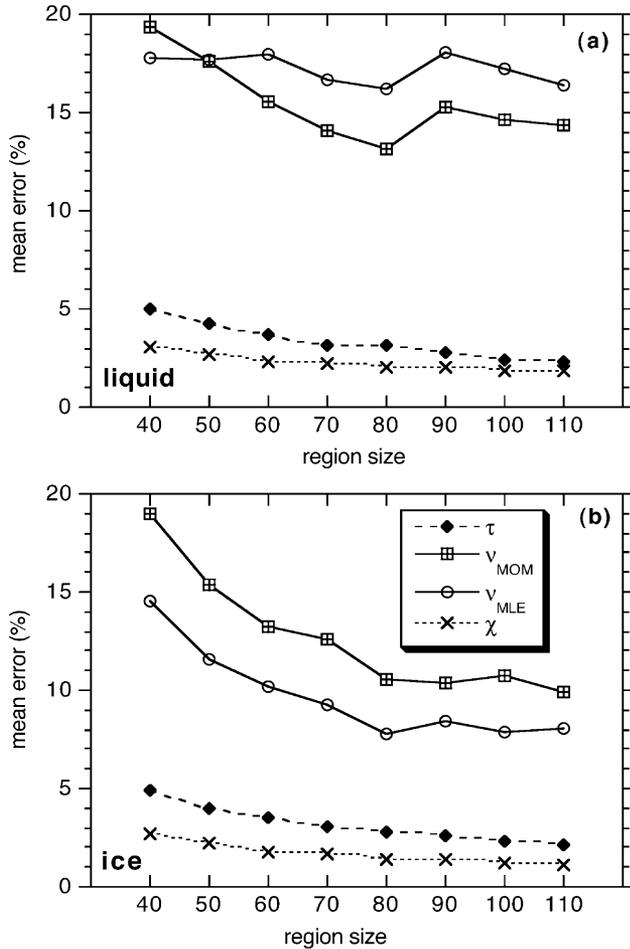


Fig. 10. As in Fig. 5(b) (save the cloud fraction), but when histograms from subsampled data are used to reconstruct the optical thickness statistics or inhomogeneity parameters. (a) Results when the Level-3 binning for liquid clouds is used. (b) Results when ice cloud binning is used.

Equations analogous to (2a) and (2b) apply for r_{eff} mean and standard deviation. The number of bins N varies according to the type of histogram, and the values used for each case will be given below. It should be underlined that the subsampling error is defined in this case as the difference between the value of the desired quantity calculated from the perfectly sampled data directly (i.e., *not* from histograms constructed from perfectly sampled data) and the value derived from (2) and histograms built from data subsampled every fifth pixel along both spatial directions.

Fig. 10 shows τ results and is the counterpart of Fig. 5(b). Panel (a) is for calculations using MODIS Level-3 binning for liquid clouds ($N = 45$ bins) and panel (b) is for calculations with ice cloud binning ($N = 30$ bins). Both histograms extend up to a value of 100 for τ , but the width of the bins is different (the ice histograms better resolve small values of τ and use coarser binning for large values). Results for both panels of Fig. 10 look similar to the results in Fig. 5(b), except for the ν_{MLE} error with liquid cloud histogram binning, which is worse for most region sizes from its counterpart for ν_{MOM} .

Fig. 11 is for r_{eff} and is the counterpart of Fig. 9(b). Fig. 9(a) is for calculations using histogram binning for liquid clouds and the bottom is for calculations with ice cloud binning. The

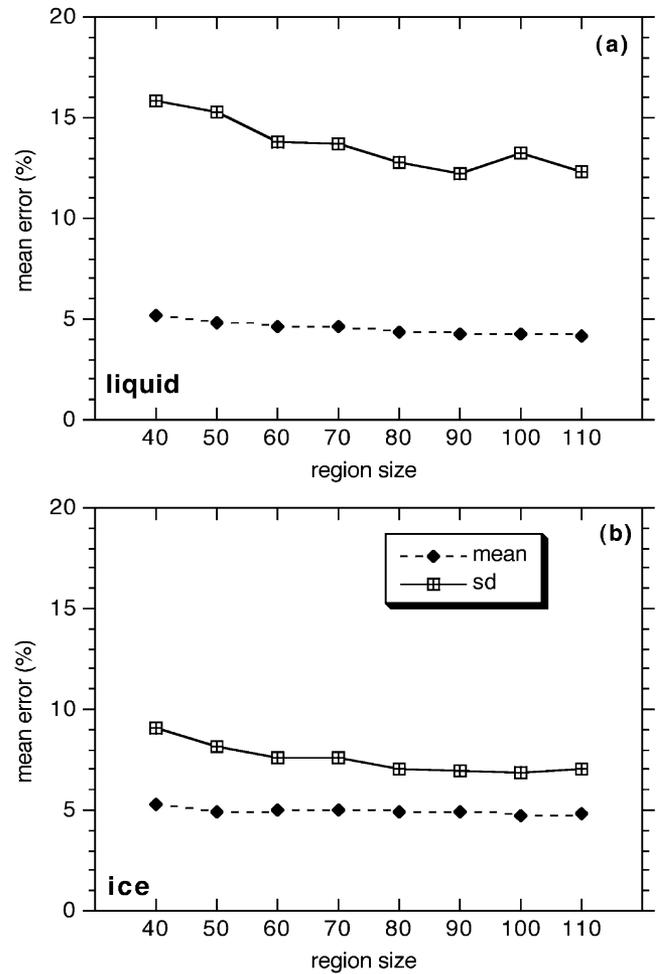


Fig. 11. As Fig. 10, but for mean and standard deviation of effective radius.

former originally uses $N = 23$ bins in the Level-3 dataset, extending from 2–30 μm , but a 24th very wide bin was added from 30–60 μm to accommodate the large particle effective radii encountered in the current dataset. The latter uses $N = 12$ histogram bins extending from 6–60 μm . Again, there is little difference from what has already been shown in Fig. 9, with the exception of the error in standard deviation when the liquid cloud histogram binning is used. This is probably the result of the coarse last bin that was arbitrarily added. Results with ice cloud binning do not seem to be much affected by omission of particle sizes below 6 μm .

In conclusion, for monthly or longer time scales, one can reconstruct τ or r_{eff} moments and τ inhomogeneity parameters from MODIS Level-3 histograms (built from data subsampled every fifth pixel) for a $1^\circ \times 1^\circ$ region, without suffering much additional subsampling error relative to the case where the moments and parameters are derived from distinct Level-3 SDSs.

VI. SUMMARY AND CONCLUSION

Cloud optical thickness and effective radius scientific datasets in the MODIS Level-3 daily, eight-day, and monthly products come from aggregation on a $1^\circ \times 1^\circ$ grid of Level-2 orbital swath data that have been subsampled every fifth pixel along both spatial directions (along and across the satellite track). The present

study examined the impact of this subsampling on cloud fraction, the mean and standard deviation of optical thickness and effective radius, as well as on parameters that convey the radiative impact of optical thickness variability. The subsampling effect was quantified as the percentage difference between perfectly sampled and subsampled results for ensembles of regions with size on the order of $1^\circ \times 1^\circ$. The perfectly sampled data come from 300 Terra and Aqua granules obtained at northern subtropics/midlatitudes for several post-2000 November days.

It was shown that Level-3 subsampling does not affect the various quantities investigated to the same degree, with second-order moments and quantities depending on second-order moments suffering greater subsampling errors, as expected. For individual regions consisting of 110×110 pixels (about $1^\circ \times 1^\circ$ around the equator), the vast majority of regions have errors within $\pm 20\%$ for mean and standard deviation of optical thickness and effective radius. Errors for cloud fraction and the inhomogeneity parameter χ are smaller, and errors for the inhomogeneity parameters ν_{MOM} and ν_{MLE} are greater (especially for ν_{MOM}). Mean errors drop dramatically when averages over a sufficient number of regions (e.g., monthly and/or zonal averages) are taken: for ensembles of 30 regions (corresponding to monthly averages) errors for most regions sizes are less than 15% for ν_{MOM} and ν_{MLE} 95% of the time, while for the other quantities they are generally below 5%. Subsampling errors seem to be mostly of random nature, but evidence was found of small but systematic underestimates for effective radius mean and standard deviation. This was traced back to systematic differences in the retrievals from different $2.1\text{-}\mu\text{m}$ band detectors: the subsampling procedure was systematically picking a pixel line (from the first two detectors) with radiatively different appearance from the other pixel lines; this pixel line is used in the actual MODIS Level-3 aggregation and subsampling algorithm. Finally, when histograms built from subsampled data with the same binning rules as in the Level-3 dataset are used to reconstruct the quantities of interest, the mean errors at monthly scales do not deteriorate significantly.

It may be worth mentioning that subsampling error analysis was also performed with the two-dimensional bounded cascade model of Cahalan [10], which offers the advantage that the properties of clouds (cloud fraction, degree of inhomogeneity, mean optical thickness) can be easily controlled. Optical thickness errors due to subsampling from MODIS largely mirrored those derived from the model clouds. The ranking of parameters according to error magnitude was the same (χ exhibited the smallest errors and ν_{MOM} the largest), the error decreased with cloud fraction and cloud homogeneity, and experienced rapid decline when averaged over ensembles of randomly generated cascade fields.

The results in this paper provide guidance to users of MODIS Level-3 cloud products on the range of errors due to subsampling they should expect and perhaps account for, in scientific work with this dataset. Although the findings do not come from a global dataset of successive satellite orbits, which would allow the Level-3 $1^\circ \times 1^\circ$ aggregation and subsampling algorithm to be better imitated (a nontrivial exercise by any means), it would probably be safe to conclude that subsampling errors should not be a serious concern for individual gridpoints of MODIS D3

(daily) data that have undergone moderate additional temporal averaging ($\sim 2\%$ error for monthly values of mean optical thickness and $\sim 3\%$ for mean effective radius), or for spatial averages such as zonal averages ($\sim 0.5\%$ and $\sim 2.5\%$ for means of optical thickness and effective radius, respectively). Still, a study of the type shown here, but with a global dataset from successive orbits spanning over a month or longer, and which would perhaps examine other SDSs as well, would give a more definitive assessment of the impact of MODIS Level-3 subsampling.

ACKNOWLEDGMENT

The author would like to thank T. Várnai, A. Marshak, S. Platnick, and two anonymous reviewers for helpful suggestions that helped improve the original manuscript.

REFERENCES

- [1] M. D. King, W. P. Menzel, Y. J. Kaufman, D. Tanré, B.-C. Gao, S. Platnick, S. A. Ackerman, L. A. Remer, R. Pincus, and P. A. Hubanks, "Cloud and aerosol properties, precipitable water, and profiles of temperature and water vapor from MODIS," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 2, pp. 442–458, Feb. 2003.
- [2] S. A. Ackerman, K. I. Strabala, W. P. Menzel, R. A. Frey, C. C. Moeller, and L. E. Gumley, "Discriminating clear sky from clouds with MODIS," *J. Geophys. Res.*, vol. 103, pp. 32 141–32 157, 1998.
- [3] S. Platnick, M. D. King, S. A. Ackerman, W. P. Menzel, B. A. Baum, J. C. Riédi, and R. A. Frey, "The MODIS cloud products: Algorithms and examples from Terra," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 2, pp. 459–473, Feb. 2003.
- [4] M. Nishihama, R. Wolfe, D. Solomon, F. Patt, J. Blanchette, A. Fleig, and E. Masuoka. MODIS Level 1A earth location: Algorithm theoretical basis document version 3.0. [Online]. Available: http://modis.gsfc.nasa.gov/data/atbd/atbd_mod28_v3.pdf.
- [5] L. Oreopoulos and R. F. Cahalan, "Cloud inhomogeneity from MODIS," *J. Clim.*, submitted for publication.
- [6] H. W. Barker, "A parameterization for computing grid-averaged solar fluxes for inhomogeneous marine boundary layer clouds, part I: Methodology and homogeneous biases," *J. Atmos. Sci.*, vol. 53, pp. 2289–2303, 1996.
- [7] L. Oreopoulos and R. Davies, "Plane parallel albedo biases from satellite observations. Part II: Parameterizations for bias removal," *J. Clim.*, vol. 11, pp. 933–944, 1998.
- [8] D. S. Wilks, *Statistical Methods in the Atmospheric Sciences*. San Diego, CA: Academic, 1995.
- [9] R. F. Cahalan, W. Ridgway, W. J. Wiscombe, T. L. Bell, and J. B. Snider, "The albedo of fractal stratocumulus clouds," *J. Atmos. Sci.*, vol. 51, pp. 2434–2455, 1994.
- [10] R. F. Cahalan, "Bounded cascade clouds: Albedo and effective thickness," in *Nonlinear Proc. Geophys.*, vol. 1, 1994, pp. 156–167.



Lazaros Oreopoulos received the B.S. degree in physics from Aristotle's University, Thessaloniki, Greece, in 1989, and the M.S. and Ph.D. degrees in atmospheric and oceanic sciences from McGill University, Montreal, QC, Canada, in 1992 and 1996, respectively.

He is currently with NASA's Goddard Space Flight Center, Greenbelt, MD, as faculty of the University of Maryland Baltimore County (UMBC) Joint Center for Earth Systems Technology since 1997. He is also affiliated with the Department of Physics at UMBC as a Research Assistant Professor. Prior to his current position he has worked for the Meteorological Service of Canada, Downsview, ON. His research interests include climate change, modeling and remote sensing of clouds, and three-dimensional radiative transfer.

Dr. Oreopoulos is a member of the American Meteorological Society and the American Geophysical Union.